CV HW4

Group 23

- M09455 Liu, Jia-Chi
- M09455 Chen, Shao-Yun
- M09414 Yu, Ruei-Lin

1. Introduction

Reconstructing a 3D scene and simultaneously obtaining the camera poses of a monocular sequence is a fundamental task in computer vision. Although it has been well-studied for decades, this task (SfM) we are able to is made an sight structure from a set of images with different view points, or equivalently from a camera in motion.

2. Principle

According to the known conditions with the structure and the motion, SfM has three types which are "pose estimation", "triangulation", and "reconstruction".

	Structure (world geometry)	Motion (camera geometry)	Measurements
Pose Estimation	known	**estimate**	2D to 3D correspondences
Triangulation	**estimate**	known	2D to 2D correspondences
Reconstruction	**estimate**	**estimate**	2D to 2D correspondences

The algorithm of SfM includes the following steps:

1. Feature Matching and Outlier rejection using RANSAC
2. Estimating Fundamental Matrix
3. Estimating Essential Matrix from Fundamental Matrix
4. Estimate Camera Pose from Essential Matrix
5. Triangulation and Reprojection to reconstruct the 3D model.
6. Bundle Adjustment

The implementation of the algorithm is in the part "Implementation procedure," but the last step is not included.

Further details, we don't provide "bundle adjustment," which refines the poses and 3D points. Instead, we focus on the geometric foundations and implementing separated tasks.

3. Implementation procedure

3.1 Feature Matching and Outlier rejection using RANSAC

Feature Matching

Firstly we use SIFT (Scale-invariant feature transform) algorithm to detect local features. The algorithm detects keypoints and descriptors about scale by finding the interest points and calculating the descriptors.

Outlier rejection by using RANSAC

After obtaining the feature points, we use brute-force matcher which selects the descriptor of one feature from one set to match with each set of other. Then we sort and prune using some distance metrics to obtain the best matches. However, there still exists outliers (false matches). In estimate a fundamental Matrix mapping one image onto the other.

3.2 Estimate Fundamental Matrix using 8-point Algorithm

We have the equation $x'^T F x = 0$
where x and x' are matched feature homogeneous coordinates of first and second image

Normalize image coordinates

We refer to the paper: Revisiting Hartley's Normalized Eight-Point Algorithm

Define $\tilde{m}$ as normalized data, $m = T\tilde{m}$
where T transform the origin data to normalized data.

$$\tilde{x} = (\frac{x_1}{\sigma_x}, \frac{x_2}{\sigma_y}, \dots, \frac{x_n}{\sigma_x})^T$$

$\tilde{x}$ is the mean value of x, σ_x is the mean value of x.

$$T = \begin{bmatrix} \frac{1}{\sigma_x} & 0 & -\frac{\bar{x}}{\sigma_x} \\ 0 & \frac{1}{\sigma_y} & -\frac{\bar{y}}{\sigma_y} \\ 0 & 0 & 1 \end{bmatrix}$$

After we normalize the coordinates, solve F

$$\begin{bmatrix} \tilde{x}'_1\tilde{x}_1 & \tilde{x}'_1\tilde{y}_1 & \tilde{x}'_1 & \tilde{y}'_1\tilde{x}_1 & \tilde{y}'_1\tilde{y}_1 & \tilde{y}'_1 & \tilde{x}_1 & \tilde{y}_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \tilde{f}_{11} \\ \tilde{f}_{12} \\ \tilde{f}_{13} \\ \tilde{f}_{21} \\ \tilde{f}_{22} \\ \tilde{f}_{23} \\ \tilde{f}_{31} \\ \tilde{f}_{32} \\ \tilde{f}_{33} \end{bmatrix} = 0$$

We perform SVD decomposition to find the null space of A and $\tilde{F}$ is the solution of $\tilde{F}$ there is F should be 0. However, due to the noise of matched pairs, rank of F may be equal to 9.

To satisfy the constraint of $\text{rank}(F) = 2$, we enforce the last singular value of F to be 0.

$$\begin{aligned} U, S, V &= \text{SVD}(\tilde{F}) \\ S &= \text{diag}(S) \\ S &= \text{zero} \end{aligned}$$

so now, we do reconstruct F

$$F = U\tilde{F}V$$

3.3 Estimate Essential Matrix from Fundamental Matrix

The essential matrix can be reconstructed from F and K. The singular value of E is not always $\Sigma = \text{diag}(1, 1, 0)$ in the noise K. So we reconstruct E using singular values $\Sigma_1 = \Sigma_2$

$$E = K^T F K$$

$$\begin{aligned} U, S, V &= \text{SVD}(E) \\ S &= \begin{bmatrix} 1 & & \\ & 1 & \\ & & 0 \end{bmatrix} \\ E &= U\Sigma V \end{aligned}$$

3.4 Estimate Camera Pose from Essential Matrix

The first camera matrix is $P_1 = [I|0]$. There are four possible choices for the second camera matrix P_2 which are:

$$\begin{aligned} W &= \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ P_2&= [UWV^T|+u_3] \\ P_2&= [UWV^T|-u_3] \\ P_2&= [UW^TV^T|+u_3] \\ P_2&= [UW^TV^T|-u_3] \end{aligned}$$

To know which P_2 is correct, use each P_2 to get 3D points using Triangulation method then pick the correct with most of 3D points in front of camera.

3.5 Triangulation method

$$x = [x \ y \ 1]^T = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad X = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$x \times PX = \begin{bmatrix} yp^{3T}X - p^{2T}X \\ p^{1T}X - xp^{3T}X \\ xp^{2T}X - yp^{1T}X \end{bmatrix} = 0$$

Same as $x' \times P'X$

$$-\begin{bmatrix} x'p'^{3T}X - p'^{1T}X \\ yp'^{3T}X - p'^{2T}X \end{bmatrix} = 0$$

The 3D point coordinate can obtain by $AX = \begin{bmatrix} yp^{3T} - p^{2T} \\ p^{1T} - xp^{3T} \\ x'p'^{3T} - p'^{1T} \\ y'p'^{3T} - p'^{2T} \end{bmatrix} X = 0$

We can solve X by computing:

```
U, S, V = SVD(A)
X = V[:,-1]
X = X/X[-1]
```

4.Experimental Results

Following the algorithm of SfM, we can obtain the following result.

4.1 Evaluation using the box images

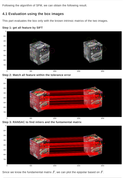
This part evaluates the box pairs with the known intrinsic matrics of the two images.

Step 1: get all feature by SIFT

Step 2: Match all feature within the tolerance error

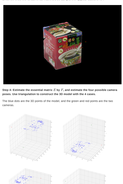
Step 3: RANSAC to find inliers and the fundamental matrix

Since we know the fundamental matrix F, we can find the epipolar lesson of F.



Step 4: Estimate the essential matrix E by F, and estimate the four possible camera poses. Use triangulation to construct the 3D model with the 4 cases.

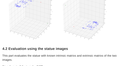
The blue dots are the 3D points of the model, and the green and red points are the two cameras.



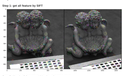
4.2 Evaluation using the statue images

This part evaluates the statue with unknown intrinsic matrices and nonmoss inspirce of the two images.

Step 1: get all feature by SIFT



Step 2: Match all feature within the tolerance error



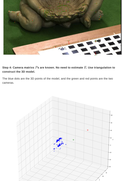
Step 3: RANSAC to find inliers and the fundamental matrix

Since we know the fundamental matrix F, we can find the epipolar lesson of F.

Blue dots represent inliers the lesson on the epipolar.

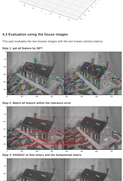
Step 4: Estimate the essential matrix E by F, and estimate the four possible camera poses. Use triangulation to construct the 3D model with the 4 cases.

The blue dots are the 3D points of the model, and the green and red points are the two cameras.



Step 5: Camera matrix P is unknown. We need to estimate C. Use triangulation to construct the 3D model.

The blue dots are the 3D points of the model, and the green and red points are the two cameras.



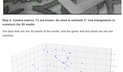
4.3 Evaluation using the House images

This part evaluates the two houses images with the two known camera matrics.

Step 1: get all feature by SIFT

Step 2: Match all feature within the tolerance error

Step 3: RANSAC to find inliers and the fundamental matrix



Since we know the fundamental matrix F, we can find the epipolar lesson of F.

The right image's epipolar line lesson on the left image.

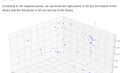
Step 4: Estimate the essential matrix E by F, and estimate the four possible camera poses. Use triangulation to construct the 3D model.

The blue dots are the 3D points of the model, and the green and red points are the two cameras.

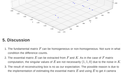
4.4 Evaluation using the library images

This part evaluates the two library images with the two known camera matrics.

Step 1: get all feature by SIFT



Step 2: Match all feature within the tolerance error



Step 3: RANSAC to find inliers and the fundamental matrix

Since we know the fundamental matrix F, we can find the epipolar lesson of F.

The right image's epipolar line lesson on the left image.

Step 4: Estimate the essential matrix E by F, and estimate the four possible camera poses. Use triangulation to construct the 3D model.

The blue dots are the 3D points of the model, and the green and red points are the two cameras.

According to the matches points, we can know the right point is 90 and the bottom of the library and the left points is 80 we can be sure of the library.

5. Discussion

1. The fundamental matrix F can be homogeneous or non-homogeneous. Not sure in what condition the difference occurs.
2. Although we know the essential matrix E is constrained from F and K, it is in the case of F metric computation, the singular values of E are not necessarily $(1, 1, 0)$ due to the noise in K.
3. The result of reconstructing box is so our expectation. The possible reason is due to the camera matrices are known and the matches points are very clear in the box images. However, we have observed that the implementation is exactly as same as the theory. As far as we know, other teams encountered the same issue that cannot reconstruct the box well, so we think it is the fault of given ground truth.

6. Conclusion

In this homework, we implement of the algorithm in lecture structure and motion from two given images. However, we can see it is robust in practical tasks. The implicit camera matrices are chosen so that they satisfy the computed fundamental matrix. The scenes are reconstructed up to projective ambiguity. Although we have implemented the algorithm, matching feed with their ambiguous. Since the structure and motion has been recovered for the entire sequence, the results can be refined through bundle adjustment.

7. Work Assignment Plan

- Liu, Jia-Chi: Code implementation, Introduction, Principle, Experimental Result, Discussion
- Chen, Shao-Yun: Bibliography, Conclusion
- Yu, Ruei-Lin: Bibliography,Code Implementation,Procedure

- Revisiting Hartley's Normalized Eight-Point Algorithm
- Scale-invariant feature transform